10.1 part 2 notes

	For	Against	No opinion	total
Democrat	50	150	50	250
Republican	125	50	25	200
Independent	15	10	25	50
total	190	210	100	500

1) To study the relationship between party affiliation and support for a balanced budget amendment, 500 registered voters were surveyed with the following results:

a) What is the row variable?

party affiliation

b) How many cells are there?

9 cells

c) What is the column variable?

support for amendment

	For	Against	No opinion	total
Democrat	50	150	50	250
Republican	125	50	25	200
Independent	15	10	25	50
total	190	210	100	500

d) What percentage of those surveyed were Democrats?

$P(D) = \frac{250}{500} = 50\%$

e) What percentage of those surveyed were FOR the amendment **and** were Republican?

$P(F \text{ and } R) = \frac{125}{500} = 25\%$

f) What percentage of those FOR the amendment are also Republicans?

$P(R \mid F) = \frac{125}{190} = 65.79\%$

g) What percentage of Independents had no opinion?

$P(\text{No} \mid I) = \frac{25}{50} = 50\%$

	For	Against	No opinion	total
Democrat	50	150	50	250
Republican	125	50	25	200
Independent	15	10	25	50
total	190	210	100	500

h) What percentage of those against the amendment were Democrats?

$P(D \mid A) = \frac{150}{210} = 71.43\%$

i) What percentage of those that had no opinion were Independents?

$P(I \mid \text{No}) = \frac{25}{100} = 25\%$

j) What percentage of those surveyed were democrats **and** had no opinion?

$P(\text{Dem and No}) = \frac{50}{500} = 10\%$

#2 is HW

Inference on 2-way tables: Chi-Square test for Association

Comparing gender of Stat students to their college location choices

	Female	Male	total
In-State	17	21	38
Out of State	9	18	27
Total	26	39	65

Question: Is there an association between the row and column variables?

- Compare… observed values in table to expected values. (affect, relationship, indep.)
- Use… χ^2 test

If we are doing a Chi-Square test, we need expected values.

Question: How do we find expected values?

Expected Cell Count = $\frac{\text{row total} \times \text{column total}}{n}$

Find the expected cell counts for the table of college choices versus gender:

OBS:

	Female	Male	total
In-State	17	21	38
Out of State	9	18	27
Total	26	39	65

EXP:

	Female	Male	total
In-State	$\frac{38 \times 26}{65} = 15.2$	$\frac{38 \times 39}{65} = 22.8$	38
Out of State	$\frac{27 \times 26}{65} = 10.8$	$\frac{27 \times 39}{65} = 16.2$	27
Total	26	39	

Chi-Square Test for Association

- Want to test…

if there is an association btw. 2 variables (categrical)

Hypotheses:

- **Ho:** there is no association btw. row var. and column var. (=)
- **Ha:** there is an association btw. _____ and _____.
  - No association = independent, no relationship, no affect

Test Statistic:

- Same!

$$\chi^2 = \sum \frac{(obs-exp)^2}{exp} =$$

*show 2 calculations, then + … =

P-Value:

- Work: $P(\chi^2 > \_\_\_\_) = \chi^2cdf(\_\_\_, E99, df)$
- df = $(\#rows-1)(\#columns-1)$

M	
E	

Conclusion:

- Same 2 sentences!
  - Reject / Fail to reject….
  - We have/do not have sufficient evidence… (re-copy Ha)

On Calculator:

- Observed → Matrix A
- run χ^2 test
  - STAT → TESTS → χ^2 test
- Expected are in Matrix B.

Conditions:

- SRS
- All expected counts ≥ 5

1. The manager of an assembly process wants to determine whether the number of defective articles manufactured depends on the day of the week the articles are produced. Using the data below, is there sufficient evidence to determine if the number of defective articles is independent of the day of the week?

Day	Mon.	Tue.	Wed.	Thur.	Fri.
Nondef.	85	90	95	95	90
Defective	15	10	5	5	10

	91	91	91	91	91
	9	9	9	9	9

Conditions:

- SRS
- Stated SRS
- all exp. counts ≥ 5
- see chart

Ho: there is no association btw. defectiveness and day of week

Ha: there is an assoc. btw. defectiveness & day of week

$$\chi^2 = \frac{(85-91)^2}{91} + \frac{(90-91)^2}{91} + \dots = 8.547$$

$P(\chi^2 > 8.547) = 0.0735$ (df=4)

We do not reject Ho b/c the p-value of 0.0735 > α=0.05.

We do not have sufficient evidence that there is an association btw. defectiveness and day of week.

2. The following table is from the July 1993 publication of *Vital and Health Statistics* from the Centers for Disease Control and Prevention/National Center for Health Statistics. The individuals in the following table have only one of the three indicated irritations. Determine if the type of irritation is independent of the age group.

Irritation	18-29	30-44	45-64	65+
Eye	440	567	349	59
Nose	924	1311	794	102
Throat	253	311	157	19

432.85	585.97	...
957.78	1296.6	...
226.37	306.44	...

Conditions:
- SRS
- Stated SRS
- all exp ≥ 5
- See table above

Ho: Irritation is independent of age.
Ha: Irritation is dependent of age.

$$\chi^2 = \frac{(440-432.85)^2}{432.85} + \frac{(567-585.97)^2}{585.97} + \dots = 13.62$$

$$P(\chi^2 > 13.62) = 0.034 \quad (df = 6)$$

We reject Ho b/c p-value of 0.034 < $\alpha = 0.05$. We have sufficient evidence that irritation is dependent on age.

3) Exp:

66.907	355.89 ...
42.033	⋮
⋮	

Conditions:
- SRS
- Assume representative
- all exp. ≥ 5
- See chart above

Ho: District is indep. of type of crime
Ha: District is dep. on type of crime

$$\chi^2 = \frac{(54-66.907)^2}{66.907} + \frac{(331-355.89)^2}{355.89} + \dots = 337.66$$

$$P(\chi^2 > 337.66) = 1.54 \times 10^{-59} \quad (df = 20)$$

We reject Ho b/c p-value of $1.54 \times 10^{-59} < \alpha = 0.01$.